SFB 649 Discussion Paper 2015-004

Efficiency of Wind Power Production and its Determinants

Simone Pieralli*
Matthias Ritter*
Martin Odening*

* Humboldt-Universität zu Berlin, Germany

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".

http://sfb649.wiwi.hu-berlin.de
ISSN 1860-5664

SFB 649, Humboldt-Universität zu Berlin
Spandauer Straße 1, D-10178 Berlin

SFB 649 ECONOMIC RISK BERLIN

Efficiency of Wind Power Production and its Determinants*

Simone Pieralli[a**], Matthias Ritter[a], Martin Odening[a]

Abstract
This article examines the efficiency of wind energy production. We quantify production losses in four wind parks across Germany for 19 wind turbines with non-convex efficiency analysis. In a second stage regression, we adapt the linear regression results of Kneip, Simar, Wilson (2014) to explain electricity losses by means of a bias-corrected truncated regression. Our results show that electricity losses amount to 27% of the maximal producible electricity. These losses can be mainly traced back to changing wind conditions while only 6 % are caused by turbine errors.

Key words: wind energy, efficiency, free disposal hull, bias correction.

JEL codes: D20, D21, Q42.

1 Introduction

Renewable energy production has experienced a rapid growth over the last two decades, and it is likely that this growth will continue. Wind energy production contributed a significant share to this expansion and has attracted institutional investors. The profitability of wind energy production is determined by generation costs, energy price, and productivity of turbines. In the past, investments in wind parks were able to attain comparably high returns on investments. In many countries, such as Germany or Spain, producers receive guaranteed prices for wind energy that range above market prices. Generation costs are also fairly stable, since operating costs are relatively low and installation costs are rather transparent. Thus, productivity turns out to be the crucial driver for the profitability of wind energy production. Productivity, in turn, heavily depends on wind conditions, i.e., wind speed and its variability, at the production site. In fact, a careful assessment of wind conditions precedes any investment in wind parks. Given the paramount importance of the wind production factor it is not surprising that much effort has been spent to develop models to predict how much of the installed capacity will actually be used during the investment period (e.g., Kusiak et al., 2013).

---

* Financial support from the German Research Foundation via CRC 649 "Economic Risk" is gratefully acknowledged.
[a] Humboldt-Universität zu Berlin, Department of Agricultural Economics, Philippstr. 13, 10115 Berlin, Germany.
** Corresponding author: simone.pieralli@agrar.hu-berlin.de, Tel.: +49 (0) 30209346847, Fax: +49 (0) 30209346841

A second determinant of productivity, however, has received little attention in the literature, namely the efficiency of wind energy production. By efficiency, we mean the distance between actual and maximal energy output given a certain level of the production factors. In the context of wind energy, the maximal producible power as a function of wind speed is depicted by a power curve. Power curves are usually calculated by turbine producers for a specific turbine type under idealized conditions[1]. Wind production in reality, however, does not take place under idealized conditions and hence, actual energy production regularly deviates from the power curve. Shortfalls can be caused, for example, by rainfall, icing, suboptimal adjustment of pitch angle and nacelle position to changing wind conditions, as well as by technical failures and scheduled maintenance. Under marginal wind conditions or a scenario of declining subsidies, these production inefficiencies can diminish the profitability of wind power plants.

A few empirical papers exist that deal with productivity and efficiency analysis of wind power generation. Homola et al. (2009) analyze wind park data in Norway and suggest a correction for power curve estimation. Ilinca (2011) reports estimated power losses due to icing conditions up to 50% of total annual production. Hughes (2012) indicates declining turbine performance due to increasing age in Denmark and the UK. Some further papers apply nonparametric methods to estimate the wind energy production frontier. Kusiak et al. (2012) use data envelopment analysis (DEA) to assess the performance of wind turbines in presence of faults. They identify turbine downtime as the major reason for power curtailment. Ibarren et al. (2014) analyze the entire process of wind energy production and include further production factors, such as land and investment cost in their DEA model. To our knowledge Carvalho et al. (2009) is the only study that applies DEA to estimate a power curve based on high frequency production data.

The objective of this paper is twofold. First, we estimate the wind energy production frontier based on production data and quantify production losses that occur in practice relative to this benchmark. In contrast to most other wind efficiency studies, we base the frontier estimation directly on high frequency production data and do not aggregate them for a single turbine or a wind park. This sheds light on the emergence of production losses over time and avoids information losses through data smoothing. Carvalho et al. (2009) pursue a similar approach. However, they use a DEA model and thus estimate efficiency by assuming a convex production technology. When doing so, one ignores the non-convex shape of a typical wind power curve and, in turn, overestimates inefficiency over some range of the production frontier below rated wind speed. To avoid this flaw of DEA, we resort to a free disposal hull (FDH) estimation of the frontier, which does not assume convexity.

Our second objective is to explain the magnitude of the observed production losses and to trace them back to factors which may or may not be under control of wind park operators. To this end, we apply a truncated regression model that accounts for biases in the regression of estimated efficiency scores in the first step of our analysis (Kneip et al., 2014). From an applied

[1] The industry standard for power curve estimation is IEC 61400-12-1 (Wächter et al., 2009; Homola et al., 2009).

perspective, our findings help improving the assessment of wind energy production under real world conditions.

In the next section, we explain in greater detail how we estimate the wind energy production frontier and derive corresponding production losses. Moreover, we present the bias correcting regression model. The methodological part is followed by a description of our data. The subsequent section presents our results, and the final section concludes.

2 Methodology

The wind kinetic energy (WK) available to be converted into electricity can be described by the following function (Hennessey, 1977; Gunturu and Schlosser, 2012):

$$\mathrm{WK} = 0.5\pi r^2 d w^3, \tag{1}$$

where r is the rotor size so that the rotor swept area is πr^2, d is the air density, and w denotes wind speed. Kinetic wind energy increases with wind speed and air density. It is important to note that according to Eq. (1), kinetic wind energy is a cubic function of wind speed. This characteristic results in a non-convex technology for wind speeds lower than rated wind speed and has implications for the estimation of the production frontier. Air density is directly proportional to air pressure and inversely proportional to air temperature. Density is higher for colder temperatures in winter and lower in summer when temperature is warmer. Air pressure causes variability in this general trend. High air pressure increases air density and low air pressure decreases density. However, only a portion of the kinetic wind energy WK can be transformed into electric power. The efficiency of this transformation process depends on various technical and managerial factors and is the subject of this study.

In general terms, the production process is characterized by a production technology which is defined as the set of all inputs (in our case: wind speed and air density) that are feasible to produce electric power:

$$T = \{w, d, e\text{: } w, d \text{ can produce } e\}, \tag{2}$$

where w is wind speed, d is air density, and e is wind electricity.

As mentioned above, wind speed is monotonically related to electrical power produced, but the rate of transformation is non-constant and increasing up to rated wind speed. However, to preserve the machine equipment from destructive centrifugal forces, the speed of rotation and thus power production are limited for wind speeds higher than rated wind speed. These features of the production technology process can be captured by a non-convex free disposal hull (FDH) for a sample of n observation points $\{w_i, d_i, e_i\}_{i=1}^{n}$:

$$\hat{T}_{\mathrm{FDH}} = \{w, d, e : w \geq w_i, d \geq d_i, e \leq e_i, \forall i = 1, \dots, n\}. \tag{3}$$

The FDH technology set creates an outer envelope of the data points included in technology T without assuming convexity. As a measure of the efficiency of the turbines in exploiting wind and air density conditions, we measure the nonparametric distance between each point and the frontier envelope. Since inputs are not controllable by producers but rather determined by nature, it is reasonable to measure distance in the direction of the outputs. We define this efficiency measure for unit (w_0, d_0, e_0) as follows:

$$\hat{\lambda}_{\mathrm{FDH}}(w_0, d_0, e_0) = \sup\{\lambda: (w_0, d_0, \lambda e_0) \in \hat{T}_{\mathrm{FDH}}\}. \tag{4}$$

An estimate $\hat{\lambda}_{\mathrm{FDH}}$ can be computed by means of a sorting algorithm that identifies all units that dominate the unit (w_0, d_0, e_0), i.e., all units that use less or equal inputs to produce equal or more output than unit (w_0, d_0, e_0).[2] Of those dominating units, the one with highest output is taken as potential producible electric power $\hat{e}_0$. The efficiency measure $\hat{\lambda}_{\mathrm{FDH}}$ is then a ratio of potential producible electric power $\hat{e}_0$ and actual electric power produced e_0. This measure of inefficiency is a conservative measure compared to convex hull measures of inefficiency, such as Data Envelopment Analysis (DEA). Given this efficiency measure, we define the production loss of electric power EL for every observation as the difference between the production potential and the electric production observed:

$$\widehat{\mathrm{EL}}_0 = \hat{\lambda}_{\mathrm{FDH}} e_0 - e_0 = \hat{e}_0 - e_0. \tag{5}$$

Previous efficiency analyses on wind energy production consider single turbines or wind parks as "decision making units" and calculate efficiency scores for these units (e.g., Ibarren et al., 2014; Iglesias et al., 2010). This kind of analysis requires to aggregate inputs and outputs to annual values. Production factors may include capital and labor. Here, we pursue a different approach. Efficiency scores are assigned to production intervals of 10 minutes length. Each observation in our sample relates the electric power produced to the average wind speed and average air density in a 10 minutes interval. Thus, we derive an efficient production function that characterizes the technology of the wind turbine under differing wind and air density conditions. This production function resembles a power curve. Other production factors than wind and air density are not considered. Pooling high frequency production data from different wind parks in an efficiency analysis makes sense if electricity is produced with the same technology (i.e., turbine type and rotor size) which is the case in our study. Productivity under this perspective can be understood as electricity output under given weather conditions. This definition differs from a productivity of a wind turbine or wind farm per unit of time, e.g., annual production. It may happen in our analysis that a wind turbine turns out to be efficient because it converts wind energy optimally into electricity, but the produced power is low due to low inputs, i.e., unfavorable wind conditions. Thus our efficiency analysis cannot support decisions on the location of wind parks or the choice among production technologies. It provides, however, useful information on the magnitude of production losses due to suboptimal utilization of wind energy.

[2] Efficiency scores λ can be alternatively determined by a mixed integer program (Deprins et al., 1984).

Such production losses may be caused by unfavorable weather conditions (apart from wind speed and air density) such as icing or by turbine faults. Though most of these factors are out of immediate managerial control, it is helpful to understand their contribution to electricity losses.

The second step of our analysis targets at explaining observed production inefficiency by regressing the estimated electricity loss $\widehat{\mathrm{EL}}$ on a set of explanatory variables $\boldsymbol{v}$. While this two-step procedure is standard for (nonparametric) efficiency analyses, it usually ignores a methodical problem that has recently been pointed out by Kneip et al. (2014). The fact that electricity losses $\widehat{\mathrm{EL}}$ result from a nonparametric estimation of the technology frontier entails problems to use them as a dependent variable in a second stage regression. Actually, the estimated efficiencies are biased measures of the true electricity losses because in a full population sample, there could be observations lying above the sample frontier. Thus, the inefficiency measures derived from the sample frontier represent a lower bound of the true ones. This bias of the regress and will spoil the estimation of the regression coefficients $\boldsymbol{\theta}$. To account for this bias, we adapt the procedure proposed by Kneip et al. (2014). A further aspect that has to be accounted for in the regression is the limited range of the calculated electricity losses: They are nonnegative and cannot exceed the maximum capacity of the turbine (2.365 MW in our sample). Thus, we apply a truncated regression model. Assuming that the latent variable, EL, is normally distributed,

$$\mathrm{EL}_i|\boldsymbol{v}_i \sim N(\boldsymbol{v}_i'\boldsymbol{\theta},\sigma^2), \tag{6}$$

the estimates of the parameters, $\widehat{\boldsymbol{\theta}}$, can be obtained by maximizing the likelihood function:

$$\mathcal{L}_1 = \prod_{\mathrm{i}=1}^{\mathrm{n}}\left(\frac{\frac{1}{\sigma}\varphi\left(\frac{\widehat{\mathrm{EL}}_i-\boldsymbol{v}_i\boldsymbol{\theta}}{\sigma}\right)}{\Phi\left(\frac{\mathrm{UB}-\boldsymbol{v}_i\boldsymbol{\theta}}{\sigma}\right)-\Phi\left(\frac{\mathrm{LB}-\boldsymbol{v}_i\boldsymbol{\theta}}{\sigma}\right)}\right) \tag{7}$$

where $\varphi(\cdot)$ is the standard normal density function and $\Phi(\cdot)$ is the cumulative standard normal density function. The arguments of the normal density and cumulative density functions derive from the conditional truncation points of the regression model. UB and LB in the argument of the cumulative distribution function are the upper bound (2,365) and the lower bound (0) of the dependent variable, respectively, and σ is the variance of the error term.

To correct the estimation bias, we split the sample in two parts and recalculate the efficiency losses $\widehat{\mathbf{EL}}_1$ and $\widehat{\mathbf{EL}}_2$, as in Kneip et al. (2014).[3] We stack these variables together to create a column vector of *n* elements $\widehat{\mathbf{EL}}^s = \begin{pmatrix}\widehat{\mathbf{EL}}_1\\ \widehat{\mathbf{EL}}_2\end{pmatrix}$. In the same way, we stack the respective *m* explanatory variables to obtain an $n \times m$ matrix $\boldsymbol{v}^s = \begin{pmatrix}\boldsymbol{v}_1\\ \boldsymbol{v}_2\end{pmatrix}$, in which the observation order is

[3] To maintain representativity of the sample in doing this estimation, we assign observations in odd positions (1, 3, 5, etc.) to the first sub-sample and in even positions (2, 4, 6, etc.) to the second sub-sample.

rearranged to match the dependent variable. Similarly to Eq. (7), we calculate a new estimator $\widehat{\widehat{\boldsymbol{\theta}}}^{s}$ in a regression of $\widehat{\mathbf{EL}}^{s}$ on $\boldsymbol{v_s}$ by maximizing the likelihood function:

$$\mathcal{L}_2 = \prod_{\mathrm{i}=1}^{\mathrm{n}} \left(\frac{\frac{1}{\sigma}\varphi\left(\frac{\widehat{\mathrm{EL}}_i^S - v_i^S \theta^S}{\sigma}\right)}{\Phi\left(\frac{\mathrm{UB} - v_i^S \theta^S}{\sigma}\right) - \Phi\left(\frac{\mathrm{LB} - v_i^S \theta^S}{\sigma}\right)} \right). \tag{8}$$

Under Theorem 5.2 in Kneip et al. (2014), the convergence result for an FDH convergence rate of $\xi \geq 1/3$ can be obtained as follows:

$$\sqrt{n}\left[\widehat{\boldsymbol{\theta}} - \left(2^{\xi} - 1\right)^{-1}\left(\widehat{\widehat{\boldsymbol{\theta}}}^{s} - \widehat{\boldsymbol{\theta}}\right) - \boldsymbol{\theta}\right] \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2 \boldsymbol{Q}) \tag{9}$$

as $n \rightarrow \infty$. In Eq. (9), $\widehat{\boldsymbol{\theta}}$ is the biased estimate and $\left(2^{\xi} - 1\right)^{-1}\left(\widehat{\widehat{\boldsymbol{\theta}}} - \widehat{\boldsymbol{\theta}}\right)$ is the bias correction, which depends on the convergence rate. The latter is defined as the inverse sum of the number of inputs (p) and outputs (q), i.e., $\xi = 1/(p+q)$. That is, the higher the number of inputs and outputs is, the higher is the bias correction.

Asymptotic confidence intervals for the vector of parameters can be determined from Eq. (9) as follows:

$$\widehat{\boldsymbol{\theta}} - \left(2^{\xi} - 1\right)^{-1}(\widehat{\widehat{\boldsymbol{\theta}}}^{s} - \widehat{\boldsymbol{\theta}}) \pm z_{1-\frac{\alpha}{2}} \hat{\sigma}_n \hat{q}_{mm} / \sqrt{n} \tag{10}$$

where $\hat{q}_{mm} = (Z'Z)^{-1}/n$ are diagonal elements of the matrix $\boldsymbol{Q}$.

3 Data and model variables

The production data used in this study refer to four wind parks, which are situated in different regions in the West, the center, and the East of Germany.[4] They consist of up to 7 turbines, which are all of the same type and capacity, namely 2.365 MW. The average produced power in Kilowatt is reported for intervals of 10 minutes from 1.7.2013 to 30.6.2014. The number of observations for all 19 turbines included in the sample sums up to 989,175. The dataset also includes observations of average mast wind for every 10 minutes interval which constitutes the first (non-controllable) input in our efficiency analysis. It should be noted that this averaging of electricity output and wind speed over a 10 minutes interval may lead to measurement errors in case of short term fluctuations of wind speed. Since wind electricity is a non-linear function of wind speed, the mean value of wind electricity differs from the wind electricity rated at mean wind speed due to Jensen's inequality. This flaw of standard methods of power curve estimation is well known and has led to modifications, for example dynamical power curve estimation (cf. Gottschall and Peinke, 2008; Homola et al., 2009). We do not adjust the measurement in our

[4] The authors thank 4inita GmbH for providing the data. The names and exact locations of the wind parks are concealed for confidentiality reasons.

efficiency analysis. However, we include wind variability as an explanatory variable for inefficiency in the regression analysis.

The second input for wind electricity production is air density. We obtain this variable from a reanalysis dataset often used in wind power analysis, namely the Modern-Era Retrospective Analysis for Research and Applications (MERRA) data provided by NASA (Rienecker et al., 2011). MERRA reanalysis data reconstruct the atmospheric state by integrating data from different sources, such as conventional and satellite data. They offer a complete worldwide grid of weather data at a spatial resolution of 1/2° latitude and 2/3° longitude (around 45 km × 54 km in Germany). We interpolate the surface air density data of the four nearest grid points weighted by their distance to each wind park.[5] The data are available at times 12:30 a.m., 1:30 a.m., 2:30 a.m., … , 11:30 p.m. for each day, which we linearly interpolate to obtain observations for every 10 minutes interval.

The summary statistics of the inputs and output variables used for the estimation of the FDH and production losses are presented in Table 1. Summary statistics of the electricity produced in all single turbines is provided in Table A.1 in the appendix.

Table 1: Summary statistics of the inputs and output

	Mean	Standard Deviation	Minimum	Maximum
Inputs				
Wind speed (m/s)	5.90	3.04	0.00	28.20
Air density (kg/m³)	1.20	0.03	1.10	1.34
Output				
Electricity produced (kW)	507.46	622.76	0.00	2,365.00
Number of observations	989,175			

The upper part of Figure 1 depicts the 10 minutes average electricity production in the four wind parks plotted against wind speed.[6] This figure provides a first impression of the range of observed wind electricity productivity given a certain level of wind speed. Apparently, the distance between highest and lowest output varies with wind speed. The highest variation in productivity can be observed at moderate wind speeds between 5 and 12 m/s whereas the productivity is rather homogeneous for calm wind conditions, as well as for observations above the rated wind speed. This is plausible because the production potential under calm wind conditions is low for technical reasons. On the other hand, the frequency of observations at

[5] The air density variable is called RHOA in the "MERRA IAU 2d surface turbulent flux diagnostics (AT1NXFLX)". More details on the MERRA products can be found in Lucchesi (2012).

[6] Note that production is measured in kW and not in kWh. This allows a direct comparison with the installed capacity but needs careful interpretation in the context of production losses. Production losses reported in kW for a 10 minutes interval can be converted to kWh through division by 6.

moderate wind speed is high (cf. Fig. 1(b)) so that the heterogeneity of outcome can be expected to be high in the presence of other production factor or measurement errors. The wind distribution in Fig. 1(b) reveals that the behavior at moderate wind speed is important for the overall efficiency of a wind turbine.

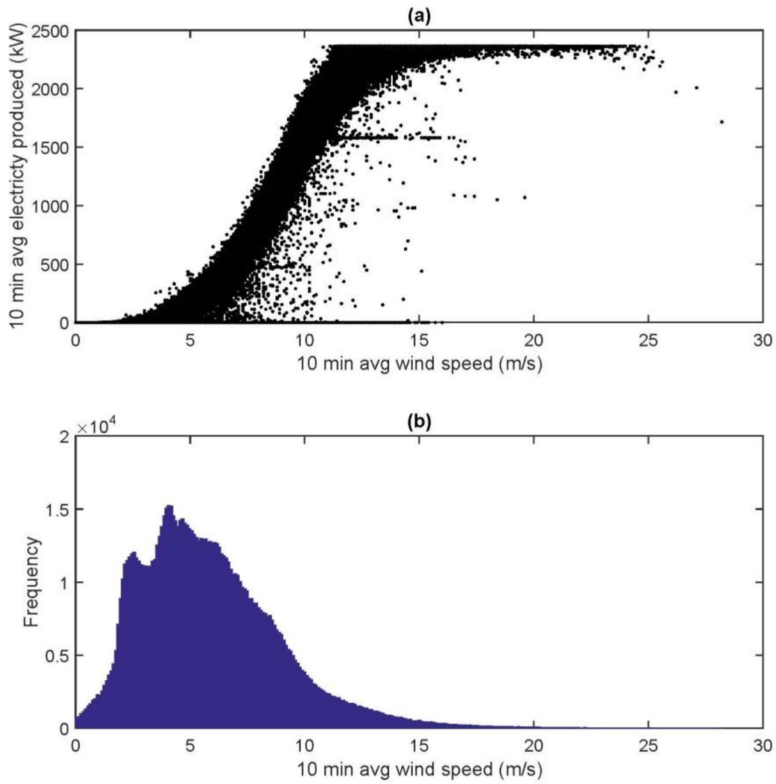

Fig. 1. (a) Power produced against wind speed, (b) Frequency of observed wind speed

Our data set provides further information that can be used to derive explanatory variables in the second stage regression. We hypothesize that variability of wind conditions, i.e., wind speed and wind direction, affect measured productivity for two reasons. First, changing wind conditions require adjustments of the turbine's operation, e.g., the rotor pitch angle or the nacelle position. These adjustments are not frictionless, will be realized with some delay, and thus will decrease power generation compared with a situation of stable wind conditions. Here, we use the range of the wind speed in a 10 minutes interval as an indicator of wind speed variability. Moreover, as mentioned above, changing wind speed will lead to measurement errors in average 10 minutes power production. We further control for the speed of adjustment of the machine to different wind speeds by considering the difference in average rotor speed between two consecutive 10 minutes intervals. Changes in the wind direction are approximated by the absolute change (either to the right or to the left) in nacelle position between two subsequent observations. The impact of this variable on productivity, however, is not clear a priori. On the one hand, higher direction variability is supposed to decrease the capability of a stable production of electric energy. On the other hand, the capability to adapt to changing wind directions can be regarded as an efficiency improving feature of the turbine. To ensure convergence of the maximum likelihood estimation, we use the cubic root of the absolute change.

Moreover, a detailed report for the turbine status is available at each time instant. This includes the occurrence of various error types as well as their starting and ending times. In the data processing, we combine this information with the 10 minutes data, i.e., we assign an error to all 10 minutes intervals between the beginning of an error and the restart of the turbine. If several errors occur concurrently, we consider only the error that occurred first. In the regression model, we consider three error categories resulting in disrupted electricity production: the presence of an ice alert on the turbine, the presence of maintenance at the turbine, and a residual error category in which all remaining errors are included. To put this in perspective: An error occurs in 30,363 observations, which is slightly more than 3% of the observations in our sample. Of these, an ice alert represents 15,077 cases, which is approximately 50% of the errors (see Fig. 2). Maintenance occurs in 5,511 cases, which are slightly more than 18% of the errors. Machine errors occur in 9,775 cases, which are the remaining 32% of the errors. All dummy variables for errors are interacted with wind speed to weigh the occurrence of an error by the wind energy.

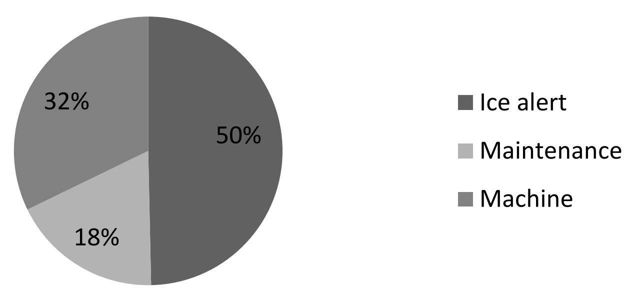

Fig. 2. Frequency of error occurrence

Finally, to account for location specific heterogeneity, we include a full set of 19 turbine dummies ($\boldsymbol{D}^{\text{turbine}}$) which capture, for example, exposition to particular wind conditions due to the position of the turbine within the corresponding wind park.

We specify the regression model of Eq (6) for the production loss EL by:

$$\text{EL} = \boldsymbol{D}^{\text{turbine}}\boldsymbol{\alpha} + \boldsymbol{Z\beta} + w\boldsymbol{H\delta} + \epsilon \tag{11}$$

where $\boldsymbol{Z} = [\text{Wind range, Nacelle position change, Rotor speed diff.}]$

and $\boldsymbol{H} = [\text{ Ice error, Maintenance error, Machine error}]$.

The summary statistics of the dependent and explanatory variables used in the second stage are presented in Table 2.

Table 2: Summary statistics of second-stage regression variables

	Mean	Standard Deviation	Minimum	Maximum
Second stage variables				
Electricity loss (kW)	190.78	150.55	0.00	2,365.00
Wind Range (m/s)	3.51	2.09	0.00	30.6
Wind speed (m/s)	5.90	3.04	0.00	28.20
Nacelle position change (o,$\sqrt[3]{\lvert x\rvert}$)	1.18	0.91	0.00	6.70
Rotor Speed Diff. (rpm)	0.0004	1.02	-17.36	16.65
Dummy Rated Speed (DR)	0.05	0.22	0	1
Number of observations	989,175			

4 Results

The resulting estimate of the FDH technology is depicted in Fig. 3. As required by the free disposability assumption on the technology, increasing amounts of both inputs are related to higher potential electricity production. The impact of the two factors on electricity production, however, is different: Low wind speed renders any air density amount unimportant for power production whereas low air density diminishes electricity production only marginally. For that reason, we focus on wind speed as the most important production factor in the subsequent figures, i.e., we focus on the power curve. Fig. 4 depicts two power curves. The broken line represents a power curve that we calculated following the industry standard IEC 61400-12-1 (Homola et al., 2009). It reflects the average produced power for wind speed bins of 0.1 m/s width where the wind speeds are adjusted to an air density of 1.225 kg/m³ and erroneous observations are excluded. The solid line is a cross-section of the FDH at the same air density of 1.225 kg/m³. It differs from the standard power curve in two ways. First, it represents an envelope of the production data instead of an average. Second, the estimation is based on all (non-filtered) observations. Apparently, both curves differ for a wide range of wind speed. The difference between the curves amounts to 183 kW per observation, which corresponds to 8% of the rated capacity. Fig. 4 shows that the distance between best performing and average points for a given wind speed is higher for wind speeds between 8 and 12 m/s. For wind speeds larger than 15 m/s, both functions converge. Note that there is a deviation between the FDH and the standard power curve for extreme wind speeds. Storm control in modern wind turbines results in a decline of power for very high wind speeds. Due to the free disposability assumption, this particular feature of wind electricity production cannot be mimicked by the FDH. In these cases, our method overestimates electricity losses. However, in our data set only 8 observations have higher wind speed than 25 m/s (cf. Fig. 1(a)).

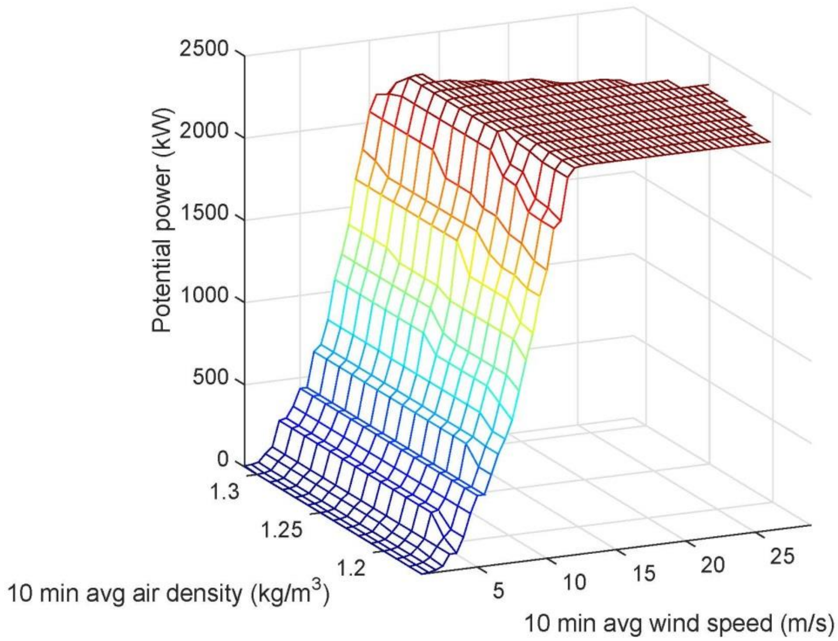

Fig. 3. Estimated free-disposal hull technology

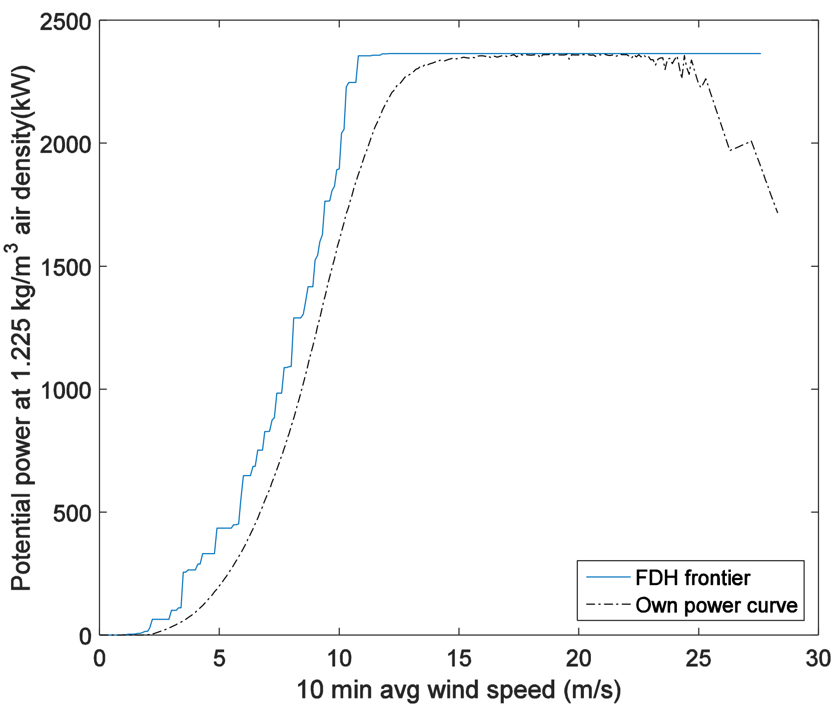

Fig. 4. Power curve vs. production frontier

The difference for every observation between potential power production represented by the FDH estimated frontier in Fig. 4 and the actual power production defines an electricity loss (EL) that is plotted against wind speed and air density in Fig. 5. Fig. 5(a) illustrates that the bulk of electricity losses occurs when approaching rated wind speed. One can further realize a bifurcation of electricity losses. While most losses decline for wind speeds larger than 10 m/s, there are a few observations that increase linearly with wind speed and then are capped at the rated power capacity. The latter losses represent faults where the turbine is partly or totally out of operation. In Fig. 5, observations with and without an error in the status code are represented by grey and black dots, respectively.

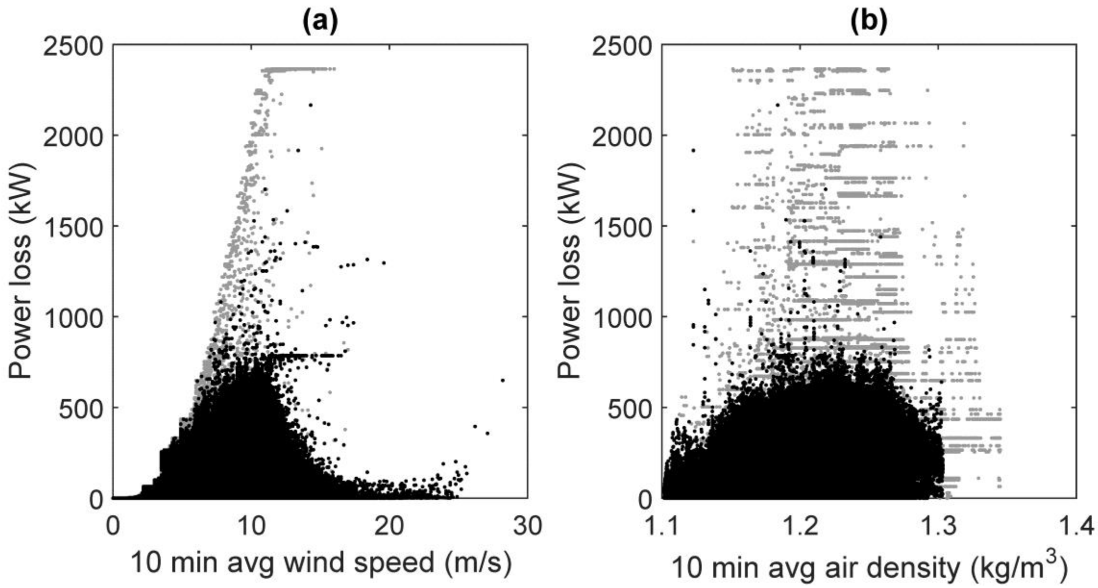

Fig. 5. (a) Power losses and wind speed; (b) power losses and air density; observations with (gray) and without (black) turbine errors

Fig. 5(b) shows that there is variability in electricity losses for different air densities. It is interesting to notice that high density is observed in presence of temperatures comparatively lower than pressure. This causes presence of ice alert and consequently a vertical cut in the figure on the right side where only observations with error are visible.

However, as mentioned for Fig. 5(a), also in the case of Fig. 5(b), the frequency of production losses is not visible because of overlapping data points. To illustrate this more clearly, Fig. 6 depicts the sum of the losses calculated in correspondence of wind speeds at 0.1 m/s intervals. Thus, the figure accounts for the severity and the frequency of the losses occurring at different wind speeds. It can be seen that the highest losses occur for wind speeds between 4 and 9 m/s whereas for lower and higher wind speeds, the cumulative losses are much smaller. This reflects the distribution of the wind speeds (Fig. 1(b)).

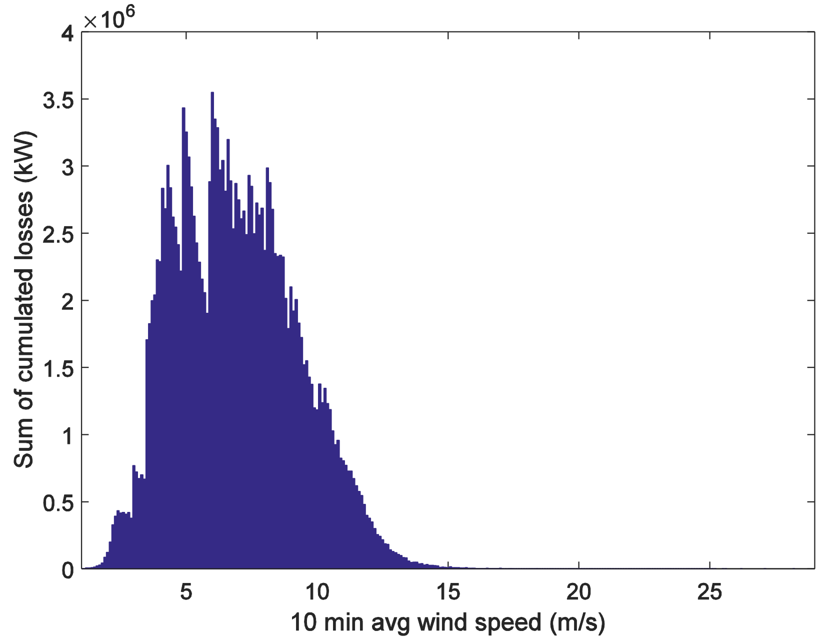

Fig. 6(a). Cumulated power losses against wind speed

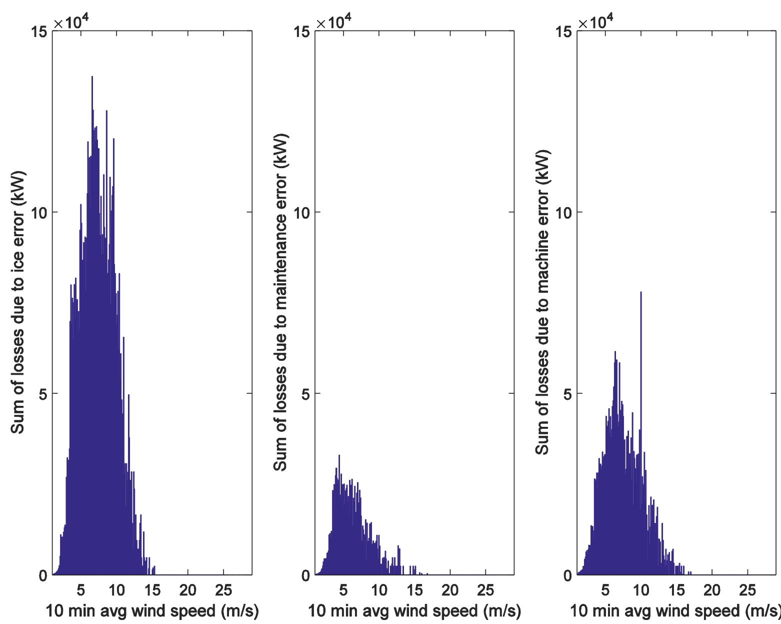

Fig. 6(b). Cumulated against wind speed for different turbine error groups

The quantitative analysis of electricity loses in Table 3 shows an average loss in 10 minutes of 191 kW associated with an average production of approximately 507 kW. That means that average losses amount to 27% of the potentially producible electricity. If one aggregates over a year, 10 minutes losses per turbine sum up to 9,930 MW which translates into 1,655 MWh.

Multiplied with an average spot price of 35.27 €/MWh attainable on the German electricity spot market, this value results in a yearly loss of 58,372 € per turbine.

Table 3: Total loss and number of observations for different error groups

	Obs	%	FDH Loss (kW)	%	FDH Loss/ obs
Ice error	15,077	1.52	7,670,313	4.06	508.74
Maintenance error	5,511	0.56	1,379,068	0.73	250.24
Machine error	9,775	0.99	3,141,902	1.66	321.42
All Errors	30,363	3.07	12,191,283	6.46	401.52
No Error	958,812	96.93	176,525,730	93.54	184.11
Total	989,175	100.00	188,717,013	100.00	190.78
Total FDH potential production			690,683,462		

Table 3 separates the losses for the three different error groups. Not surprisingly, the average loss in case of an error is much higher than without an error (401.52 kW compared to 184.11 kW). However, only 6.46% of the total losses are caused by errors because they occur only for 3.07% of the observations. Among the three error groups, ice has the biggest share with 4.06% of the total loss. The average loss in case of an ice error equals 508.74 kW, which is very similar to the overall mean of produced electricity (507.46 kW, Table 1). We conclude that ice errors lead to a total drop down of electric production and occur for all wind speeds (Fig. 6(b)). The average loss of 250.24 kW in case of a maintenance error is much lower and not far from the average loss without an error (184.11 kW), which indicates that maintenance—if possible—is conducted in periods where little losses are expected, i.e., during light wind conditions (Fig. 6(b)). The average loss of 321.42 kW in case of a machine error is higher than the maintenance error but lower than the losses incurred during icing conditions.

The results of the second stage regressions are presented in Table 4. We ran three models of different complexity. Model 1 is a parsimonious model that includes the variability of wind speed and direction, error dummies, and turbine dummies as regressors. All coefficients are highly significant which is not surprising given the large number of observations. The coefficient of the wind speed range is negative. This can be explained by the concavity of some parts of the production frontier: If the wind speed is constantly at or below the cut-in speed, no electricity will be generated. On the other hand, a mean preserving spread of the wind speed will result in some positive output during the observed 10 minutes interval and thus reduce observed losses. The coefficient of the wind range should be interpreted in conjunction with the effect of a change of the rotor speed since the latter is a response to the former and both variables are not independent from each other. We find that a higher change of the rotor speed over time is positively related to electricity losses, which basically reflects the physical energy required to overcome inertia of rotor blades. Surprisingly, adjustments of the nacelle position have an opposite effect, i.e., they reduce production losses. As expected, all error dummies interacted with wind speed are positively related to power losses and the magnitude of the coefficients confirms the earlier analysis, i.e., icing errors have the highest impact among the three considered error types. Turbine dummy variables show a positive sign which shows a positive average loss depending on turbine specific wind conditions and position in the park.

The fit of the predicted to the observed losses in Model 1—calculated via Pearson's ρ^2—is, however, modest (ρ^2=0.28). Though the fit or predictive power is not the key issue in stage-two regression analyses of technical efficiency, we modify the base regression model to attain a better fit to our data. There is a controversial discussion on whether input factors that are used in the nonparametric estimation of the production frontier should also enter the second stage regression (Simar and Wilson, 2007; Kneip et al., 2014). This amounts to the violation of a separability condition on the production technology. Disregarding these potential theoretical flaws, we include the wind speed in the set of regressors in Model 2. This can be justified by the fact that potential power output varies with wind speed and thus the impact on wind electricity losses is different depending on the speed. In fact, the sign of the wind speed is positive while the signs of all other variables remain the same as in the base model. The inclusion of this variable increases the ρ^2 considerably from 0.28 to 0.41.

Inspection of Fig. 5(a) suggests that the impact of factors on realized production losses depends on whether the wind speed is above or below rated wind speed. To distinguish these two regimes, we interact in Model 3 all variables with a dummy variable (DR) indicating a wind speed at or above the rated wind speed of 11.5 m/s.[7] Separating these two wind regimes increases the ρ^2 further to 0.63. Again, the sign of the coefficients remain the same as before apart from the rotor speed variable which is now negative. The coefficients associated with the dummy variable for high wind speed regime capture the difference between the regression coefficients for wind speed above rated speed relative to the coefficients below rated wind speed. This difference is positive for the wind variability. Varying wind speeds around a high mean value typically represent gusty and turbulent wind conditions which may disturb power production, e.g., via storm control, while the maximal capacity cannot be further increased. For the error variables, this means that their impact on power losses is even stronger than in the low wind regime. Only the effect of the wind speed decreased under the high wind regime, which is plausible since losses are lower after rated wind speed. Comparing the turbine dummies reveals considerable differences of the productivity at different locations. The best performing turbine in our sample (A1) is located in a wind park in a position that it is free from obstacles in the most recurring wind directions. None of the other turbines in this study have the same privilege.

[7] This wind speed is the empirically recovered rated wind speed in our sample, that is the first occurrence of an average production of 2,365 kW in a 10 minutes interval is at 11.5 m/s.

Table 4: Second stage regression results

Dependent variable: Electricity loss (kW)	Truncated regression model					
	Model 1		Model 2		Model 3	
Wind Range	-1.101	***	-11.792	***	-0.332	***
Nacelle position change($\sqrt[3]{\lvert x \rvert}$)	-12.361	***	-8.549	***	-6.817	***
Ice error*Wind speed	76.346	***	90.358	***	82.119	***
Maintenance error*Wind speed	25.436	***	45.661	***	38.673	***
Machine error*Wind speed	58.217	***	72.398	***	58.849	***
Rotor speed diff.	5.124	***	2.454	***	-0.589	***
Wind speed			16.568	***	34.923	***
Wind Range *DR					12.506	***
Nacelle position change *DR					75.149	***
Ice error*Wind speed*DR					154.512	***
Maint. error*Wind speed*DR					64.504	***
Machine error*Wind speed*DR					75.706	***
Rot. speed diff. *DR					14.978	**
Wind speed*DR					-49.938	***
A1	434.145	***	287.684	***	124.344	***
B1	437.149	***	308.448	***	140.722	***
B2	431.775	***	305.745	***	139.863	***
B3	430.911	***	309.061	***	147.127	***
B4	434.334	***	299.450	***	129.991	***
B5	451.763	***	308.846	***	124.861	***
B6	438.381	***	301.511	***	131.402	***
B7	429.294	***	305.203	***	144.920	***
C1	424.015	***	296.618	***	139.792	***
C2	420.881	***	293.061	***	138.054	***
C3	423.832	***	301.221	***	145.960	***
C4	428.083	***	301.607	***	142.611	***
C5	417.072	***	286.462	***	136.712	***
C6	420.409	***	289.573	***	134.045	***
D1	422.769	***	289.764	***	133.259	***
D2	419.365	***	291.877	***	137.401	***
D3	431.860	***	290.234	***	124.822	***
D4	419.091	***	294.521	***	136.140	***
D5	411.033	***	283.066	***	138.753	***
Pearson's ρ^2	0.249		0.409		0.631	

** and *** denote statistical significance at the 5 and 1 percent level, respectively.

5 Conclusions

This article analyzes the productivity and efficiency of wind electricity generation under real world conditions. Based on a sample of 19 wind turbines located in different wind parks in

Germany, we calculate an efficient production frontier that represents the maximum producible electricity given a certain level of wind speed and air density. This view on productivity considers wind as production factor and results in a production frontier that is similar to a power curve. With this frontier at hand, we can quantify electricity losses that have been realized under various wind conditions compared with this benchmark. The production frontier dominates the standard power curve by construction since the latter represents average production under some idealized conditions. We find that the difference between the frontier and the standard power curve is on average 183 kW. Thus, the standard power curve can be regarded as a conservative estimate of electricity output conditional on the exogenous wind input. Our results show that production inefficiencies sum up to a loss of 27% of the producible electricity. In a subsequent step, we decompose these production losses. It is noteworthy that turbine errors are responsible for only 6% of the production losses though they often cause a complete stop of production. The reason is that only 3% of the observations in our sample were affected by errors. Hence, the production process can be regarded as technically quite reliable. Among the losses caused by turbine errors, icing had the highest impact. We employ a regression model to explain the occurrence of production losses in greater detail. It turns out that beside turbine errors, changing wind conditions, i.e., variations of wind speed and direction, affect efficiency of electricity production. Moreover, turbine specific effects exist which likely can be traced back to the position of a turbine within a wind park.

From a methodological point of view, this paper represents—to the knowledge of the authors—the first empirical estimation of efficiency with high frequency wind electricity production data using non-convex analysis methods. Moreover, it is the first estimation of bias-corrected explanations of efficiency in production by adapting the linear regression convergence results in Kneip et al. (2014) to the truncated regression case.

It is not straightforward to draw immediate managerial conclusions from the empirical findings since most factors are not controllable, at least in the short run. Weather conditions are entirely exogenous and stochastic. Nevertheless, it is important to understand how vulnerable wind electricity production is with regard to these conditions. This knowledge emphasizes that weather conditions should be inspected carefully prior to the location decisions of wind parks. Moreover, our analysis is helpful to assess the trade-off between the benefits from increasing the distance between single turbines in a wind park and higher costs for land acquisition. Finally, our results highlight the gains arising from flexible adjustments of wind turbines to changing weather conditions. Technical progress, such as anti-icing and de-icing systems, targets at increasing this flexibility. In fact, anti-icing and de-icing systems, even if costly, could prevent important losses in the medium to long-run.

6 References

Carvalho, A., Gonzalez, M.C., Costa, P., Martins, A. (2009): Issues on Performance of Wind Systems derived from Exploitation Data. Industrial Electronics, 2009. IECON '09. 35th Annual Conference of IEEE: 3599-3604.

Deprins, D., L. Simar and H. Tulkens. (1984): Measuring Labor Efficiency in Post Offices. In M. Marchand, P. Pestieau and H. Tulkens (eds.), The Performance of Public Enterprises: Concepts and Measurements. Amsterdam: North Holland.

Gunturu and C. A. Schlosser (2012): Characterization of wind power resource in the United States. Atmospheric Chemistry and Physics, 12(20): 9687-9702.

Hennessey Jr., J. (1977): Some Aspects of Wind Power Statistics. Journal of Applied Meteorology 16: 119-128.

Homola, M.C., Byström, J., Nicklasson, P.J., Sundsbo, P.A. (2009): An improved method for wind power estimation. Working Paper, Narvik University College, Norway, 1-13.

Hughes, G. (2012): The Performance of Wind Farms in the United Kingdom and Denmark. Renewable Energy Foundation, London.

Ibarren, D., Vazquez-Rowe, I., Rugani, B., Benetto, E. (2014): On the feasibility of using emergy analysis as a source of benchmarking criteria through data envelopment analysis: A case study for wind energy. Energy 67: 527-537.

Ilinca, A., (2011). Analysis and Mitigation of Icing Effects on Wind Turbines, Wind Turbines, Dr. Ibrahim AlBahadly (Ed.), ISBN: 978-953-307-221-0, InTech, Available from: http://www.intechopen.com/books/windturbines/analysis-and-mitigation-of-icing-effects-on-wind-turbines.

Kneip, A., Simar, L., Wilson, P.W. (2014): When Bias Kills the Variance: Central Limit Theorems for DEA and FDH Efficiency Scores. Econometric Theory, FirstView http://dx.doi.org/10.1017/S0266466614000413.

Kusiak, A., Verma, A., Wei, X. (2012): Wind turbine capacity data from SCADA. Wind Systems Magazine. September 2012: 36-39.

Kusiak, A., Zhang, Z., Verma, A. (2013): Prediction, operations, and condition monitoring in wind energy. Energy 60: 1-12.

Lucchesi, R. (2012). File specification for MERRA products (version 2.3). GMAO Office Note No. 1. http://gmao.gsfc.nasa.gov/pubs/docs/Lucchesi528.pdf.

Rienecker, M. M., Suarez, M. J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., Bosilovich, M. G., Schubert, S. D., Takacs, L., Kim, G.-K., Bloom, S., Chen, J., Collins, D., Conaty, A., da Silva, A., Gu, W., Joiner, J., Koster, R. D., Lucchesi, R., Molod, A., Owens, T., Pawson, S., Pegion, P., Redder, C. R., Reichle, R., Robertson, F. R., Ruddick, A. G., Sienkiewicz, M., and Woollen, J. (2011). MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications. Journal of Climate 24(14).

Simar, L. and Wilson. P.W. (2007): Estimation and inference in two-sage semi-parametric models of production processes. Journal of Econometrics 136: 31-64.

Wächter, M., Gottschall, J., Rettenmeier, A., Peinke, J. (2009): Power curve estimation using LIDAR measurements. In: Proceedings of the EWEA European Wind Energy Conference, Marseille, France, 16–19 March 2009. Vol.1: 4473-4477.

7 Appendix

Table A.1: Summary statistics of electric power production of individual turbines

	Mean	Standard Deviation	Minimum	Maximum
Produced power (kW)	507.46	622.76	0	2,365
By turbine:				
A1	659.69	699.27	0	2,365
B1	406.14	496.44	0	2,365
B2	465.19	554.64	0	2,365
B3	438.52	541.08	0	2,365
B4	445.77	534.40	0	2,365
B5	455.51	540.41	0	2,365
B6	451.54	542.09	0	2,365
B7	460.88	553.64	0	2,365
C1	540.56	645.49	0	2,365
C2	542.04	659.38	0	2,365
C3	516.36	641.55	0	2,365
C4	532.76	632.25	0	2,365
C5	604.09	709.68	0	2,365
C6	553.62	666.26	0	2,365
D1	534.79	647.18	0	2,365
D2	504.27	643.73	0	2,364
D3	462.71	621.11	0	2,364
D4	485.51	644.01	0	2,364
D5	582.42	732.41	0	2,365
Number of observations	989,175			

SFB 649 Discussion Paper Series 2015

For a complete list of Discussion Papers published by the SFB 649, please visit http://sfb649.wiwi.hu-berlin.de.

001 "Pricing Kernel Modeling" by Denis Belomestny, Shujie Ma and Wolfgang Karl Härdle, January 2015.
002 "Estimating the Value of Urban Green Space: A hedonic Pricing Analysis of the Housing Market in Cologne, Germany" by Jens Kolbe and Henry Wüstemann, January 2015.
003 "Identifying Berlin's land value map using Adaptive Weights Smoothing" by Jens Kolbe, Rainer Schulz, Martin Wersing and Axel Werwatz, January 2015.
004 "Efficiency of Wind Power Production and its Determinants" by Simone Pieralli, Matthias Ritter and Martin Odening, January 2015.

SFB 649, Spandauer Straße 1, D-10178 Berlin
http://sfb649.wiwi.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".